VOLUME-8, ISSUE-1, JANUARY-2019 • PRINT ISSN No 2277 - 8160

Original Research Paper

Forensic Science

NGM SELECT STR MARKERS IN A POPULATION SAMPLE OF UKRAINIAN IMMIGRANTS LIVING IN PORTUGAL

Bruno Monteiro Faculty of Sciences of the University of Porto, FCUP, Porto, Portugal

MJ Prata* Faculty of Sciences of the University of Porto, FCUP, Porto, Portugal and Institute of Molecular Pathology and Immunology of the University of Porto, IPATIMUP, Porto, Portugal *c

L Pontes* National Institute of Legal Medicine and Forensic Sciences, INMLCF, Portugal *Corresponding Author

ABSTRACT In this work, a sample of Ukrainians living in Portugal was characterized for the autosomal markers included in the *NGM Select kit*, a *kit* primarily used in forensic casework analysis. The estimated values of heterozygosity, combined power of discrimination (PDc) and combined probability of paternity exclusion (PEc), demonstrated the high level of discrimination between Ukrainians afforded by this *kit*, thereby making it very informative and consequently an important asset in forensic routine. Interpopulation comparisons showed significant genetic differentiation between the sample of Ukrainians and populations from non-European regions. Within Europe, low population structure, modestly related with geography, was detected with this *kit*. These findings pinpoint the prospective use of the *NGM Select kit* mainly in population genetic studies addressing the genetic differentiation between continental populations.

KEYWORDS : Short Tandem Repeats; Population data; Ukrainians; Intrapopulation and interpopulation comparisons

Introduction

Since 2000, and especially during the first decade, the number of Ukrainian immigrants living in Portugal increased as much as 213%. In 2011, Ukrainian immigrants were the third biggest community of immigrants living in Portugal, constituting 9% (33 790 individuals) of the total immigrants recorded in the Portuguese 2011 Census [1]. Thus, the aim of this work was to characterize the Ukrainian community residing in Portugal for 16 autosomal STRs included in the *NGM Select kit*, to evaluate whether population structure could represent a concern in the forensic setting. Viewing that, population comparisons were further performed to evaluate how this sample of Ukrainians complied with the patterns of genetic diversity reported for other populations.

Material/Methods

Extraction: DNA was extracted from buccal cell swabs by the Chelex method [2].

Quantification: DNA quantification was achieved by real-time PCR using the QuantifilerTM Human DNA Quantification Kit (AB) and the ABI PRISM® Sequence Detection System (AB).

PCR Amplification: The extracted DNA was amplified using the NGM Select Amplification *Kit* (Applied Biosystem, Foster City, CA) according to the manufacturer's recommendations, in a GeneAmp 9700 PCR system (Applied Biosystem, Foster City, CA).

Electrophoresis: PCR products were analyzed by capillary electrophoresis in an Applied Biosystems 3500 Genetic analyzer (Foster City, CA). All samples were analyzed at least twice and only results doubly confirmed were considered.

Data analysis: Allele assignment and designation was carried out by comparison with reference sequenced ladders and was performed with the GeneMapper ID-X v1.4 analysis software. Arlequin software v.3.5.2.2 [3] was used to calculate allele frequencies, *locus by locus* population pairwise genetic distance (F_{ST}), expected heterozygosity (He), observed heterozygosity (Ho) and to test departures from Hardy–Weinberg equilibrium.

Power of Discrimination and Matching Probability were calculated using GenAlEx v6.503 software package [4]. Probability of Exclusion, combined probability of paternity exclusion, polymorphic information content and paternity index were obtained with Exel using the formulas described in the literature [5].

Average pairwise genetic distance (FST) over all *loci* was calculated with the Poptrew software [6] and submitted to Multidimensional Scaling Analysis using the R v3.3.2 software [7], and then the yielded coordinates were used to construct the corresponding graphs in Excel.

Populations used: The sample studied consisted in a total of 72 Ukrainian immigrants living in Portugal who gave informed consent to participate in the study under anonymized identification. The number of individuals analyzed is relatively small, however it is representative of the total of Ukrainian immigrants living in Portugal, since the number of individuals corresponds to 0,32% of the total population of 10.6 million Portuguese [1].

For the comparative analysis, frequency data regarding the *loci* included in the *NGM Select kit* was recruited from the literature for samples from Portugal [8] and 15 additional worldwide populations namely Lithuania, Albania, Slovenia, Iraq, Turkey [9], Bosnia and Herzegovina, Germany [10], Netherlands [11], Spain [12], Denmark, Greenland [13], Kazakhstan [14] and Somalia [13] as well as from the AmaXhosa, a bantu-speaking group of South Africa [15], and the Han from the Fujian province of China [16].

Locus by *locus* F_{ST} comparisons for Ukraine with the other populations were performed. The F_{ST} values and respective p vales, the matrix of F_{ST} and $(\delta\mu)^2$ distances for all the populations and the MDS graph for F_{ST} data were also obtained (data not shown).

Results and discussion

Allele frequencies, heterozygosities and forensic statistic parameters for the 16 STR *loci* investigated in the Ukrainians living in Portugal are shown in Figure 1.

Allele	Loci															
	D10S1248	vWA	D16S539	D2S1338	D8S1179	D21S11	D18S51	D22S1045	D19S433	TH01	FGA	D2S441	D3S1358	D1S1656	D12S391	SE33
6	-	-	-	-	-	-	-	-	-	0.21528	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-	-	0.12500	-	-	-	-	-	-
8	-	-	0.01389	-	0.00694	-	-	-	-	0.13889	-	-	-	-	-	-
9	-	-	0.12500	-	0.01389	-	-	-	-	0.21528	-	-	-	-	-	-
9.3	-	-	-	-	-	-	-	-	-	0.30556	-	-	-	-	-	-
10	-	-	0.03472	-	0.07639	-	0.00694	-	-	-	-	0.25694	-	-	-	-
11	-	-	0.27778	-	0.07639	-	0.00694	0.19444	-	-	-	0.26389	-	0.06250	-	-
11.3	-	-	-	-	-	-	-	-	-	-	-	0.09722	-	-	-	-
12	0.02778	-	0.34028	-	0.15278	-	0.05556	0.02083	0.05556	-	-	0.02778	-	0.15278	-	0.00694
12.2	-	-	-	-	-	-	-	-	0.00694	-	-	-	-	-	-	-
13	0.18750	-	0.18056	-	0.38194	-	0.16667	0.00694	0.26389	-	-	0.05556	-	0.08333	-	0.02778
13.2	-	-	-	-	-	-	-	-	0.05556	-	-	-	-	-	-	-
14	0.36111	0.11111	0.02778	-	0.17361	-	0.13889	0.04861	0.29167	-	-	0.27083	0.13194	0.10417	0.00694	0.02083
14.2	-	-	-	-	-	-	-	-	0.02778	-	-	-	-	-	-	-
15	0.26389	0.12500	-	0.00694	0.09028	-	0.15972	0.31250	0.15972	-	0.00694	0.02083	0.23611	0.10417	0.01389	0.04167
15.2	-	-	-	-	-	-	-	-	0.02778	-	-	-	-	-	-	-
15.3	-	-	-	-	-	-	-	-	-	-	-	-	-	0.02778	-	-
16	0.15278	0.14583	-	0.04167	0.02083	-	0.23611	0.30556	0.04861	-	-	0.00694	0.23611	0.15972	0.04167	0.09028
16.2	-	-	-	-	-	-	-	-	0.04167	-	-	-	-	-	-	-
16.3	-	-	-	-	-	-	-	-	-	-	-	-	-	0.05556	-	-
17	0.00694	0.25694	-	0.28472	0.00694	-	0.09722	0.06944	0.00694	-	-	-	0.20139	0.01389	0.11806	0.03472
17.2	-	-	-	-	-	-	-	-	0.00694	-	-	-	-	-	-	-
17.3	-	-	-	-	-	-	-	-	-	-	-	-	-	0.15972	0.00694	-
18	-	0.25694	-	0.06944	-	-	0.04861	0.04167	0.00694	-	0.02083	-	0.18056	0.01389	0.27083	0.07639
18.3	-	-	-	-	-	-	-	-	-	-	-	-	-	0.06250	0.01389	-
19	-	0.09028	-	0.08333	-	-	0.05556	-	-	-	0.08333	-	0.01389	-	0.09722	0.06250
19.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.02083	-
20	-	0.01389	-	0.20833	-	-	0.01389	-	-	-	0.18056	-	-	-	0.12500	0.09722
20.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.02083
21	-	-	-	0.03472	-	-	0.00694	-	-	-	0.19444	-	-	-	0.09028	0.03472
21.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.02083
22	-	-	-	0.02778	-	-	0.00694	-	-	-	0.18750	-	-	-	0.09722	0.01389
22.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.02083
23	-	-	-	0.10417	-	-	-	-	-	-	0.10417	-	-	-	0.07639	0.00694
23.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.02083
24	-	-	-	0.08333	-	-	-	-	-	-	0.12500	-	-	-	0.01389	-
24.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.00694
25	-	-	-	0.04167	-	-	-	-	-	-	0.06250	-	-	-	0.00694	-
25.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.02083
26	-	-	-	0.01389	-	-	-	-	-	-	0.03472	-	-	-	-	-
26.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.07639
27	-	-	-	-	-	0.02083	-	-	-	-	-	-	-	-	-	-
27.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.09722
28	-	-	-	-	-	0.18750	-	-	-	-	-	-	-	-	-	-
28.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.03472
29	-	-	-	-	-	0.17361	-	-	-	-	-	-	-	-	-	-
29.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.06250
30	-	-	-	-	-	0.22222	-	-	-	-	-	-	-	-	-	-
30.2	-	-	-	-	-	0.06944	-	-	-	-	-	-	-	-	-	0.04861
31	-	-	-	-	-	0.04167	-	-	-	-	-	-	-	-	-	-
31.2	-	-	-	-	-	0.09028	-	-	-	-	-	-	-	-	-	0.01389
32	-	-	-	-	-	0.02778	-	-	-	-	-	-	-	-	-	-
32.2	-	-	-	-	-	0.08333	-	-	-	-	-	-	-	-	-	0.00694
33.2	-	-	-	-	-	0.08333	-	-	-	-	-	-	-	-	-	0.02083
34	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.00694
34.2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.00694
Ho	0.70833	0.81944	0.75000	0.83333	0.83333	0.88889	0.75000	0.83333	0.86111	0.81944	0.84722	0.83333	0.84722	0.91667	0.86111	0.94444
He	0.74582	0.81605	0.76195	0.84615	0.78555	0.86150	0.85927	0.76709	0.81342	0.78448	0.86140	0.78263	0.80332	0.89112	0.86733	0.94784
P	0.60766	0.62188	0.95300	0.22110	0.71997	0.75687	0.04210*	0.38836	0.27809	0.62214	0.45746	0.25520	0.99579	0.80110	0.46041	0.10571
MP	0.11090	0.06228	0.09776	0.04232	0.07385	0.03698	0.03803	0.09379	0.06114	0.08403	0.03765	0.08459	0.07259	0.02431	0.03237	0.00654
PD	0.88910	0.93772	0.90224	0.95768	0.92615	0.96302	0.96197	0.90621	0.93886	0.91597	0.96235	0.91541	0.92741	0.97569	0.96763	0.99346
PE	0.44126	0.63560	0.50977	0.66229	0.66229	0.77278	0.50977	0.66229	0.71688	0.63560	0.68939	0.66229	0.68939	0.82959	0.71688	0.88677
PIC	0.69701	0.75589	0.65153	0.82347	0.75461	0.83941	0.83679	0.72473	0.78359	0.74383	0.83867	0.74225	0.76606	0.87386	0.84817	0.93817
PI	1.71427	2.76916	2.00000	2.99994	2.9994	3.50005	2.00000	2.99994	3.59997	2.76916	3.27268	2.99994	3.27268	6.00024	3.59997	8.99928
	PDc = 0.99999999999999999999829836					**PEc** = 0.99999999213598			**Ho média** = 0.83420 +/- 0.06011					**He média** = 0.82468 +/- 0.05465		

Ho: observed heterozygosity. He: expected heterozygosity. P: p value for Hardy-Weinberg equilibrium test. MP: matching probability. PD: discrimination power. PE: exclusion probability. PIC: polymorphic information content. PI: paternity index. PDc: combined power of discrimination. PEc: combined probability of exclusion.
** after Bonferroni correction (p = 0.003125)*

Figure 1: Forensic and populational parameters obtained for a sample of Ukrainians living in Portugal (n=72)

No significant deviations from the Hardy–Weinberg equilibrium were found, even for the locus D18S51 (p = 0.040), once after applying the Bonferroni´s correction for multiple testing (adjusted p = 0.0031) the associated p value lost statistical significance.

The He values for each *locus* were all high, ranging from 0.7458 in the D10S1248 locus to as much as 0.9478 in the SE33. Accordingly, the PIC values were also very elevated, usually ≥0.7, except the D16S539 (0.65153), which was also one of the least heterozygous *locus*. The paternity index (PI) varied from 1.7 (not surprisingly at D10S1248) to 9 (at SE33, as predictable), while the 16 STR *loci* afforded a combined paternity index of 1.1×10^8. The combined power of discrimination (PDc) and the combined probability of paternity exclusion (PEc) for all the loci are 0.999999999999999999998 and 0.999999992, respectively, illustrating the usefulness of this 16-STR multiplex for kinship analysis in the population sample scrutinized.

Here we present 5 additional STR loci (D10S1248, D22S1045, D2S441, D1S1656 and D12S391) from the European Standard Set STR (ESS), who were posteriorly added in 2009. These markers are designated of mini-STR, because of their amplicon size of 70-180 bp. They provide competitive ability when compared to SNPs or INDELs, for the analysis of degraded samples. The inclusion of these markers maximize the sensibility of degraded DNA analysis, improve the discrimination power and the quality of the results [17].

These five *loci* have not been described in the literature until know, and their inclusion in this study contributes significantly to the observed high discrimination power. Two of these *loci* (D1S1656 and D12S391) have high vales of He and PD, being the most informative *loci*, after SE33. These *loci* are, inclusively, more informative than the ones, previously described in the literature (D21S11, FGA, D18S51 and D2S1338), as very informative [18]. The loci previously described in the literature, as very informative, were also very informative in our analysis. The SE33 locus was already described in the literature [18]; however, we present also the allelic frequencies and forensic parameters of the Ukrainian sample for the SE33 locus, who is the most informative *locus* of this kit.

The computed l*ocus*-by-*locus* F_{ST} distances involving the Ukrainians living in Portugal and other populations, revealed the clear predominance of significant differences between Ukrainians and population from regions far from Europe, like Greenland, Asia (Kazakhstan, China) and Africa (South Africa, Somalia). Distances of Ukrainians comparatively to Middle Eastern populations (from Iraq and Turkey) were higher than compared to European populations, but still none of them reached statistical significance. The differentiation of Ukrainians relatively to other Europeans was very low and by large non-significant, including comparatively to the Portuguese, a result that alleviates the concerns that detection of substructure could pose from the forensic perspective.

These findings translate the ability of the set of markers contained in *NGM Select kit* to discriminate sharply between continental populations, ability that, however, is overtly insufficient to discriminate between European populations, likely because overall Europe is characterized by high level of genetic homogeneity [19]. The Multidimensional Scaling graph based on the matrix of F_{ST} value for a total of 16 populations (Fig 2), captured well this wide pattern. They exhibit the clear clustering of European populations apart from non-European ones, while also reveal that within Europe the distribution of population diversity is barely structured and only modestly related to geography. Previously, the correspondence between genetic and geographic distances in Europe was clearly observed in studies based on both dense number of *loci* and dense geographic sampling [19]. Thus, is quite understandable that the resolution afforded by the limited set of markers here studied was not enough to apprehend the broad structure of special variation of diversity among European populations.

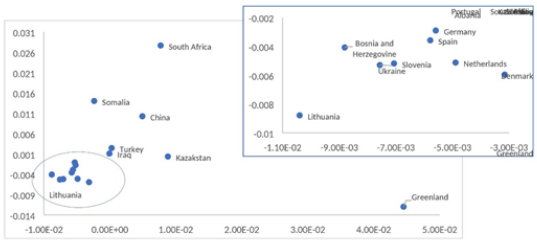

Figure 2: Multidimensional scaling (MDS) graph for FST values obtained with the matrix of the genetic distance (for 15 of the 16 markers) for the populations compared.

REFERENCES

[1] Instituto Nacional de Estatística, S. P. (2011). A população estrangeira em Portugal.

[2] Walsh, P. S., Metzger, D. A., & Higushi, R. (2013). Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. BioTechniques 10(4): 506-13 (April 1991). Biotechniques, 54(3), 134-139.

[3] Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour, 10(3), 564-567.

[4] Peakall, R., & Smouse, P. E. (2012). GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. Bioinformatics, 28(19), 2537-2539.

[5] Huston, A. K. (1998). Statistical Analysis of STR Data1(3), 14-15. Delghandi, M., Branting, B. J., & Jorgensen, L. (2001). Evaluation of a Quadruplex Short Tandem Repeat System (HUMVWA31/A, HUMD11S554, HUMAPOI1, and HUMACTBP2 loci) for Forensic Identity Testing, Confident Typing of Complex Alleles, and Population Databases. Croatian Medical Journal, 42(1), 33–44.

[6] Takezaki, N., Nei, M., & Tamura, K. (2014). POPTREEW: web version of POPTREE for constructing population trees from allele frequency data and computing some other quantities. Mol Biol Evol, 31(6), 1622-1624.

[7] Team, R. C. (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from http://www.R-project.org/

[8] Pontes, M. L., & Pinheiro, M. F. (2012). Population data of the AmpFISTR(R) NGM STR loci in a North of Portugal sample. Forensic Sci Int Genet, 6(5), e127-128.

[9] Poulsen, L., Tomas, C., . . . Morling, N. (2016). NGMSElect and Investigator((R)) Argus X-12 analysis in population samples from Albania, Iraq, Lithuania, Slovenia, and Turkey. Forensic Sci Int Genet, 22, 110-112.

[10] Welch, L. A., . . . Bastisch, I. (2012). European Network of Forensic Science Institutes (ENFSI): Evaluation of new commercial STR multiplexes that include the European Standard Set (ESS) of markers. Forensic Sci Int Genet, 6(6), 819-826.

[11] Westen, A. A., . . . de Knijff, P. (2014). Comparing six commercial autosomal STR kits in a large Dutch population sample. Forensic Science International: Genetics, 10, 55-63.

[12] Garcia, O., . . . Yurrebaso, I. (2012). Population genetic data and concordance study for the kits Identifiler, NGM, PowerPlex ESX 17 System and Investigator ESSplex in Spain. Forensic Sci Int Genet, 6(2), e78-79.

[13] Tomas, C., . . . Morling, N. (2014). Concordance study and population frequencies for 16 autosomal STRs analyzed with PowerPlex(R) ESI 17 and AmpFlSTR(R) NGM SElect in Somalis, Danes and Greenlanders. Forensic Sci Int Genet, 11, e18-21.

[14] Zhang, H., . . . Li, S. (2016). Population genetic analysis of the GlobalFiler STR loci in 748 individuals from the Kazakh population of Xinjiang in northwest China. Int J Legal Med, 130(5), 1187-1189.

[15] Ristow, P. G., Cloete, K. W., & D'Amato, M. E. (2016). GlobalFiler((R)) Express DNA amplification kit in South Africa: Extracting the past from the present. Forensic Sci Int Genet, 24, 194-201.

[16] Li, B., Ge, J., Wu, F., Ye, L., Budowle, B., & Chen, Y. (2013). Population genetic analyses of the STR loci of the AmpFlSTR NGM SElect kit for Han population in Fujian Province, China. Int J Legal Med, 127(2), 345-346.

[17] Schneider, P. M. (2009). Expansion of the European Standard Set of DNA database loci-the current situation. Profiles in DNA, 12(1), 6-7.

[18] Delghandi, M., Branting, B. J., Jorgensen, L. (2001). Evaluation of a Quadruplex Short Tandem Repeat System (HUMVWA31/A, HUMD11S554, HUMAPOI1, and HUMACTBP2 loci) for Forensic Identity Testing, Confident Typing of Complex Alleles, and Population Databases. Croatian Medical Journal, 42, 33-44.

[19] Seldin, M. F., . . . Gregersen, P. K. (2006). European Population Substructure: Clustering of Northern and Southern Populations. PLoS Genetics, 2(9), e143.